J. Vis. Commun. Image R. 38 (2016) 55–64

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Background light ray modeling for change detection ☆

Atsushi Shimada [a,*], Hajime Nagahara [b], Rin-ichiro Taniguchi [b]

[a] Faculty of Arts and Science, Kyushu University, 819-0395, Japan
[b] Faculty of Information Science and Electrical Engineering, Kyushu University, Japan

ARTICLE INFO

Article history:
Received 21 August 2015
Revised 28 December 2015
Accepted 24 February 2016
Available online 27 February 2016

Keywords:
Change detection
Light field
Background modeling
Spatio-temporal consistency

ABSTRACT

This paper is an extension of the work that was originally reported in Shimada et al. (2013). This paper proposes a change detection method based on spatio-temporal light ray consistency. The proposed method introduces light field sensing, which is used to generate an arbitrary in-focus plane. Change detection is performed in a surveillance scene, where the background region can be filtered out by an out-focusing process. This approach resolves a longstanding issue in background modeling-based object detection, which often suffers from false positives in the background regions. To realize this new change detection method, a new feature representation, called the local ray pattern (LRP), is introduced. The LRP evaluates the spatial consistency of the light rays, and this plays an important role in distinguishing whether the light rays come from the in-focus plane or elsewhere. A combination of the LRP and Gaussian mixture model (GMM)-based background modeling realizes change detection in the in-focus plane. Experimental results demonstrate the proposed method's effectiveness and its applicability to video surveillance.

1. Introduction

Change detection based on background modeling is often used in visual surveillance applications. Changes in the image signal are detected by creating a statistical model of the observed pixel values. Change should be caused by the detection of targets such as people walking or moving cars. However, background changes, including waving branches and cast shadows, are also factors in that they cause false positive detection. While several studies have focused on effective background modeling methods [2,3], there is, to date, no radical or effective solution to this problem.

Our study proposes a new sensing strategy that enables the system to exclude the background regions from the space of interest at the imaging stage.[1] Throughout this paper, we refer to the "space of interest" as the in-focus area in which target detection is performed. In contrast to the in-focus area, the background region is called the out-focus area. The proposed sensing strategy can create an in-focus area with an arbitrary shape, as shown in Fig. 1. Therefore, the background region (e.g., the tree in Fig. 1) is only captured as a blur. To realize this imaging strategy, we introduce a light field camera to capture the light rays from the scene, and apply a digital refocusing technique to generate the in- and out-focus areas. Change detection is also performed by processing the captured light rays from the viewpoint of spatio-temporal light field consistency.

The main contributions of this paper are

1. Configuration of in-focus areas with arbitrary shapes using a light field camera.
2. Detection of change in the in-focus areas based on light ray processing.

We implemented a new surveillance system using the ProFUSION 25 light field camera, and performed experiments using real scenes in which traditional change detection methodologies would suffer from false positive detection problems.

2. Related work

The light field camera, which was originally proposed for image-based rendering by the graphics community, has been used for a wide variety of visualization applications, including computer imaging through virtual apertures, 3D graphics, and digital refocusing [5]. Early light field cameras, such as the Stanford multi-camera array [6], were large and expensive systems. However, the latest light field cameras, which consist of a micro-lens array located between the sensor and the main lens, are becoming both

☆ This paper has been recommended for acceptance by Kiyoharu Aizawa.

* Corresponding author.
E-mail address: atsushi@ait.kyushu-u.ac.jp (A. Shimada).

[1] We have intentionally used the term "space of interest" rather than "region of interest (ROI)" because ROI has an implicit 2D meaning.

http://dx.doi.org/10.1016/j.jvcir.2016.02.013
1047-3203/

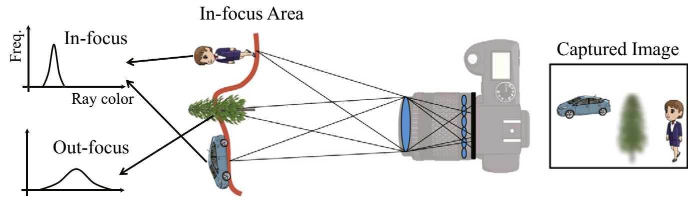

Fig. 1. Proposed in-focus area: only the areas in which the target objects appear are focused, while other areas are captured as blurs. Distribution of light ray colors: if the light rays originate from the in-focus area, there is a small variance; otherwise, the variance is larger. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inexpensive and compact [7], and are thus becoming available in the product market. The light field camera is therefore becoming a popular input device for computer vision applications. Most recently, light field cameras have been used to solve difficult computer vision and pattern recognition problems, such as transparent object recognition [8].

Single-view cameras have mainly been used in visual surveillance applications, in which background modeling-based change detection is a fundamental technique [2–4]. Change detection is performed on a 2D image plane, and traditional approaches are therefore frequently affected by changes in the background. The method proposed in this paper uses a light field camera, and represents the detection field as a 3D space, which allows the background region to be filtered out of the detection field.

A 3D intrusion detection system using multiple cameras was previously proposed [9]. In that system, an arbitrary 3D volumetric restricted area was generated by multiple cameras surrounding the target area. Intrusion detection was then achieved by computing the intersection between the object and the sensing plane, which acts as the boundary of the restricted area. The method proposed in this paper also generates a restricted area for change detection, but the camera arrangement and the detection algorithm used here both differ greatly from those of the previous work. The proposed method introduces a digital refocusing technique to generate the restricted area, and change detection is then achieved via light ray processing.

Stereo or multiple camera systems have also been considered as 3D approaches [10–12]. These approaches are effective for operation under severe conditions, such as illumination changes or occlusions. One of the main advantages of using multiple view approaches is that they can use depth information (i.e., geometric information) to reduce the background changes. Our proposed method is also a type of multiple camera system (a stereo vision is a subset of it) which captures multiple light rays simultaneously. Our motivation of introducing a light filed camera is to captures the light rays with higher density. As we will report in the experimental result section, ordinary cameras using two or three light rays fails to detect really changes in the scene because of less accuracy of matching between rays. Richer light ray information enables to detect the changes more accurately.

A combination of a camera and a depth sensor is one practical way of using depth information. A background subtraction technique based on color and depth information was proposed previously [13]. In this approach, an active sensor (the Microsoft Kinect) is used to measure the depth. We believe that such an active sensor is only applicable to indoor scenes, and would not be suitable for outdoor applications. In contrast, our approach only requires passive sensors, which are, of course, suitable for outdoor use.

A strategy of generating an arbitrarily focused image by fusing multiple differently focused images can provide similar effect with our proposed strategy in terms of focusing on particular depth and filtering out others [14]. This approach captures a couple of images with different focus in advance, then generate a fused image. In contrast, our proposed strategy captures multiple light rays, then generate an arbitrarily focused image directly from the light rays. This strategy is well known as digital refocusing [5] as mentioned above.

The original concept behind our approach has already been introduced in [1]. The original paper explained the main part of the concept and provided minimal experimental results. The paper did not demonstrate the effectiveness of light field sensing for the solution of change detection issues in sufficient depth. This paper therefore discusses the experimental results in detail, including a quantitative evaluation that involves varying the number of cameras, the light ray densities, and the scenes, and also discusses the computation costs. These results demonstrate the effectiveness of the proposed approach, which was not discussed in [1].

3. Overview of the proposed method

First, we set up an in-focus area in which the change detection process is to be performed. As shown in Fig. 1, the in-focus area does not have to be a 2D plane. For example, it is possible to intentionally set up an "out-focus area" in which background changes occur frequently (refer to Section 6 for an actual situation). To configure the in- and out-focus areas, we use a digital refocusing technique and a light field camera.

Next, we evaluate two factors: the spatial and temporal consistencies of the observed light rays. The appearance of a Lambertian object is view-independent, and all rays in a light field that correspond to the same point in a scene must thus have similar intensity values. In other words, the intensity variance associated with light rays emerging from a Lambertian scene point is low, as shown in Fig. 1. We use this characteristic to evaluate the spatial consistency. Also, the temporal consistency of the light rays is modeled using a Gaussian mixture background model, which is often used in change detection applications. With regard to the Lambertian reflectance assumption, most of the target objects, such as people, cars, and baggage, which are likely to be detected in real situations, have Lambertian surfaces even if they would have partially specular surfaces. While the complete shape of the target is not segmented (i.e., partially undetected pixels exist), the proposed approach can detect most parts of the target, and this is sufficient for subsequent processing such as tracking and event detection. The same discussion also applies for occlusion scenes (i.e., where part of the target is occluded by another object).

Finally, the evaluation results for the spatial and temporal consistencies are integrated to determine foreground masks that denote the change detection result. A Markov random field-based approach is then introduced to assign foreground/background labels to all the light rays.

4. In-focus area configuration

We use a 4D-ray representation of the light field image $L(s,t,u,v)$ that is determined by the intersection of the camera plane (s,t) and the slant of the ray (u,v). We use a commercial light